High Performance Layout-Friendly 64-Bit Priority Encoder Utilizing Parallel Priority Look-Ahead

Khaled M. Ali, Hassan Mostafa

Abstract—A novel high-performance priority encoder design using full custom approach is presented. The new encoder design implementation provides both high- and low-priority functionalities with scalable design structure through a parallel look-ahead structure. The prefixing architecture is applied to minimize the critical path propagation delay which results in increasing the maximum operating frequency. The proposed encoder shows significant improvement in terms of speed, and regulatory in building higher order encoders. Simulation results are conducted for different encoder inputs through TSMC 130nm CMOS technology.

I. INTRODUCTION

PRIORITY Encoders (PE) which generate an output stream based on the highest prioritised input stream are utilized extensively in digital and computer systems as microprocessors[5][6][9]. Due to the high demand for faster computer systems so it is a key point to start to optimize the speed of (PE) to improve the whole response of the system. At the same time, the overall power dissipation is required to be low and simultaneously, it is required to achieve small chip area to be concurrent with scaling phenomena. The mechanism of priority encoder is based on the priority token which passed consecutively from the highest priority bit to the lowest priority bit as the high priority bits start to lose their priority first. So the maximum operating frequency of a priority encoder is totally dependent on the propagation delay which means that the critical path delay of a priority encoder is proportional to the number of inputs. As a result, the design of a CMOS priority encoder is usually valid for few number of inputs, in the range of 4 bits or 8 bits. Then, when this encoder blocks are realised using CMOS technology, the largest signal propagation path consists of a series connection of either nMOS or pMOS transistors. Hence, higher order priority encoders are constructed by using smaller size priority encoder blocks in cascade structure based on a look-ahead scheme similar to that of adders/subtractors. Few novel lookahead schemes have been presented by researchers [1][2]. In addition, the parallel priority look-ahead structure discussed in [2] appears to be perfect, which enables high-speed and low power consumption priority encoder to be implemented.

In this paper, a modification of the first stage in the parallel look-ahead structure is presented by replacing the Static CMOS OR gate by Dynamic CMOS OR gate [14] which reduces the critical delay of the OR cell by 83%. This results in reducing the total critical delay to be reduced by a factor of 53%, and reduce the total number of transistors from 647 to 599 which reduce the total chip area. These results are discussed in Section II.

The remaining part of this paper is organized as follows. The proposed 8-bit CMOS priority encoder design is discussed in Section III. The proposed architecture of the parallel priority look-ahead is discussed in section IV. The simulation method and design metrics estimated for different 64-bit (PE) blocks are given in Section V. Finally, the conclusions are drawn in Section VI.

II. PROPOSED 8-BIT DYNAMIC CMOS OR CELL

Dynamic logic[10][11] (clocked logic) is a design methodology in combinatory logic circuits, particularly those implemented in MOS technology. Dynamic logic circuits are usually faster than static counterparts, and require less total area. The overall power consumption of dynamic logic may be higher or lower depending on various tradeoffs.

Dynamic logic is distinguished from so-called static logic in that dynamic logic uses a clock signal in its implementation of combinational logic circuits. The usual use of a clock signal is to synchronize transitions in sequential logic circuits. For most implementations of combinational logic, a clock signal is not even needed[13].

In dynamic logic, there is no mechanism to drive the output high or low. In the most common version of this concept, the output is driven high or low during distinct parts of the clock cycle. During the time intervals when the output is not being actively driven, its impedance causes it to maintain a level within some tolerance range of the driven level.

Consider a 64-bit PE look-ahead structure that uses 8-input OR gates to test if there is a logic-1 in each set of eight inputs, based on a concept of divide and conquer. That is, for the first 8 bits,

\[ OR_0 = D_0 + D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7 \]  \( (1) \)

If all the inputs are low, the output of OR gate is low and the priority token is passed to the next 8-bit cell. The look-ahead signals \( LA_0 \) to \( LA_7 \) can be generated for each 8 bits at the same time.

Fig. 1 shows the dynamic logic circuit of 8-input OR gate which requires two phases. The first phase, when Clock is low, is called the precharge phase and the second phase, when Clock is high, is called the evaluation phase. In the precharge phase, the output is driven high unconditionally (no matter the values of the inputs). The capacitor, which represents the load capacitance of this gate, becomes charged. Because the transistor at the bottom is turned off, it is impossible for the output to be driven low during this phase.
During the evaluation phase, Clock is high. If all input bits are low, the output will be pulled low. Otherwise, the output stays high.

This modification improves the overall delay of PE by 53% as actually we eliminate the low to high propagation delay of this cell so we dropped the OR gate propagation delay by a factor of 83% also this mechanism improves the chip area as the total number of transistors will be reduced. However, the total power consumption will be increased due to the circuit natural behavior[12].

III. PROPOSED 8-BIT CMOS PRIORITY ENCODER DESIGN

The basic equations which cover the proposed 8-bit priority encoder shown in Fig. 2 are specified below; where \( P_0 \) to \( P_7 \) represent the primary inputs, while \( EP_0 \) to \( EP_7 \) represent the primary outputs. It is noted that the primary outputs are allowed to be valid based on the input stream and their priority assignment at the rising-edge of the clock (CLOCK) in addition to the look-ahead input signal (LA) is active high.

\[
\begin{align*}
EP_0 &= (P_0) \\
EP_1 &= (P_1)(P_0) \\
EP_2 &= (P_2)(P_0)(P_1) \\
EP_3 &= (P_3)(P_2)(P_1)(P_0) \\
EP_4 &= (P_4)(P_3)(P_2)(P_1)(P_0) \\
EP_5 &= (P_5)(P_4)(P_3)(P_2)(P_1)(P_0) \\
EP_6 &= (P_6)(P_5)(P_4)(P_3)(P_2)(P_1)(P_0) \\
\end{align*}
\]

The 8-bit CMOS priority encoder introduced above synthesizes the equations mentioned earlier in the way of sharing common logic and based on the domino logic style. In Fig. 1, the pMOS transistors marked as pc1 to pc8 act as precharge transistors which ON during the falling-edge of CLOCK that way pull-up the primary outputs \( EP_0 \) to \( EP_7 \) to be high. When CLOCK changes from a low-to-high transition (rising-edge) and provided LA is active high (logic ‘1’), pMOS transistors pc1 to pc8 are OFF when the evaluation phase started. Now a collection of the nMOS transistors ev1 to ev15 might be ON based on the values of primary inputs. From the equations listed above, it can be understood that input \( P_0 \) (and eventually \( EP_0 \)) is considered the highest priority along the input bits of the 8-bit priority encoder cell. The order of priority goes down sequentially from \( P_0 \) to \( P_7 \); also for outputs \( EP_0 \) to \( EP_7 \). However, it is noted that priority assignment for primary inputs (outputs) is ideally based on user choice.

During the precharge phase, CLOCK signal is active low (logic ‘0’); so transistors pc1 to pc8 are ON and the primary outputs \( EP_0 \) to \( EP_7 \) go high again. We now describe two mechanisms during the evaluate phase when CLOCK undergoes a rising edge transition (and eventually becomes active high), with input signal LA also still logic high state. These two mechanisms are representative of typical circuit operation

- \( P_0 \) is pulled-up high : In this case, transistor ev1 is ON and \( EP_0 \) is driven to logic high – this occurs whatever the data values of other primary inputs. Minimum data path latency occurs for this mechanism as bits \( P_0 \) and \( EP_0 \) assumed to be the highest priority.
- \( P_7 \) is pulled-up high and \( P_0 \) to \( P_6 \) are low : In this case, nMOS transistors ev2, ev4, ev6, ev8, ev10, ev12, ev14 and ev15 are ON leading to high state for \( EP_7 \). Otherwise nMOS transistors ev1, ev3, ev5, ev7, ev9, ev11 and ev13 are OFF. Maximum data path delay is encountered during this mechanism as \( P_7 \) and \( EP_7 \) are assumed to be the lowest priority.

IV. THE PARALLEL PRIORITY LOOK-AHEAD ARCHITECTURE

Lately Hang et al. [1] proposed a Priority Look-ahead technique named Multilevel Folding architecture. Compared to the Multilevel Look-ahead architecture [3], which can reduce the priority propagation critical delay for a N-bit PE to the order of \( O(N) \), the Multilevel Folding architecture can reduce it to the order of \( O(\log_2 N) \).

However, the Multilevel Folding method is complex and the connections between PE cells are not easily done. Due to the harder look-ahead connections, complicating layout and
testing, especially when the number of bits becomes large. Parallel Priority Look-ahead architecture shows that PE can be simplified in structure to improve performance.

Consider a 64-bit PE look-ahead structure that uses 8-input OR gates to test if there is a (logic “1”) in each set of eight inputs, based on a concept of divide and conquer. That is, for the first 8 bits as mentioned in Equation 1.

\[ \begin{align*}
    L_{A0} &= OR_0 \\
    L_{A1} &= OR_1 \cdot OR_0 \\
    L_{A2} &= OR_2 \cdot OR_1 \cdot OR_0 \\
    L_{A3} &= OR_3 \cdot OR_2 \cdot OR_1 \cdot OR_0 \\
    L_{A4} &= OR_4 \cdot OR_3 \cdot OR_2 \cdot OR_1 \cdot OR_0 \\
    L_{A5} &= OR_5 \cdot OR_4 \cdot OR_3 \cdot OR_2 \cdot OR_1 \cdot OR_0 \\
    L_{A6} &= OR_6 \cdot OR_5 \cdot OR_4 \cdot OR_3 \cdot OR_2 \cdot OR_1 \cdot OR_0 \\
    L_{A7} &= OR_7 \cdot OR_6 \cdot OR_5 \cdot OR_4 \cdot OR_3 \cdot OR_2 \cdot OR_1 \cdot OR_0
\end{align*} \]

(3)

From the previous equations it is clear that the look-ahead logic might be generated using an additional PE cell, with \( OR_0 - OR_7 \) as inputs and \( L_{A0} - L_{A7} \) as outputs. Fig. 3 shows a circuit that utilizes this concept to implement a Parallel Priority Look-ahead architecture. The OR gates are designed by using Dynamic CMOS logic as mentioned above in Section II.

All the PEs are implemented using the new power optimized 8-bit PE cells mentioned above in Section III.

The Parallel Priority Look-ahead architecture achieve several advantages. First, the look-ahead signals \( L_{A0} - L_{A7} \) are generated simultaneously. So, the lower significance PE cells do not have to wait for the look-ahead signals from the higher significance PE cells. For a 64-bit PE, the total gate delay consists of three gate delays: one OR gate, one AND gate in the ‘look-ahead’ PE, and one AND gate in the ‘data’ PE. Second, the look-ahead signal routing is much more regular than the Multilevel Folding architecture, which makes it possible to estimate the signal propagation delay along the wire due to the parasitic capacitance/resistance. Then optimization is performed during layout by decreasing the total area. Third, this new architecture divides the data processing into two stages, the OR gate stage and the PE stage. This makes a pipelined structure possible, which makes it scalable to use this design to build higher order design.

A 64-bit PE with a latch-based two-stage pipeline structure is illustrated in Fig. 3. The outputs of the OR gates are latched by N-CMOS latches, which used to restore data during clock change. The two stages are clocked by two nonoverlapping clock phases; when the OR stage is in the evaluation phase, the PE stage is in the precharge phase. In this phase, \( OR_0 - OR_7 \) are generated and latched on the next clock edge when the PE stage enters the evaluation phase. The ‘look-ahead’ PE reads \( OR_0 - OR_7 \) and changes one of the outputs in \( L_{A0} - L_{A7} \) determining which ‘data’ PE works and generates the final EP stream.

V. SIMULATION AND RESULTS

A 64-bit PE utilizing the new Delay-Optimized 8-bit OR cell and the Parallel Priority Look-ahead architecture in a latch-based two-stage pipelined structure and a 64-bit PE with conventional three-level look-ahead 8-bit cell and three-level folding technique (conventional design) are designed in 1.2V, 130nm TSMC CMOS technology. Post-layout simulation with all parasitics included shows the propagation delay through the new PE is 0.87ns with a 10MHz clock, as shown in Fig. 4.
4. This delay represents worst case test when the input bits stream $D_0 : D_{63}$ is 0x00 00 00 00 00 00 00 01 which requires the maximum delay to check all inputs, and the C. PE [1] has a 1.8ns delay. Total chip area is also calculated and this design exhibit lower layout area compared to C. PE [1]. Fig. 5 illustrates the chip Floor plan and Fig. 6 shows the chip Layout with total area of 1600 $\mu$m$^2$.

<table>
<thead>
<tr>
<th>Design metrics</th>
<th>C. PE</th>
<th>new PE</th>
<th>C. PE</th>
<th>new PE</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Transistors</td>
<td>647</td>
<td>599</td>
<td>647</td>
<td>599</td>
<td>Decreased by 8%</td>
</tr>
<tr>
<td>area($\mu$m$^2$)</td>
<td>2400</td>
<td>1600</td>
<td>2400</td>
<td>1600</td>
<td>Decreased by 33.3%</td>
</tr>
<tr>
<td>Delay(ns)</td>
<td>1.8</td>
<td>0.87</td>
<td>1.68</td>
<td>0.77</td>
<td>Decreased by 53%</td>
</tr>
<tr>
<td>Power($\mu$W)</td>
<td>0.79</td>
<td>0.82</td>
<td>4.205</td>
<td>16.17</td>
<td>Increased by 26%</td>
</tr>
</tbody>
</table>

Table I

Comparison of design parameters between the conventional PE and the modified one at different operating frequency

From the previous table it is clear that the power consumption is increased by a factor of 4 due to the usage of Dynamic CMOS logic OR cell as mentioned before in Section II. In addition, the new design maximum operating frequency might be extended to the range of GHz but the conventional design is limited to the range of MHz.

REFERENCES


VI. CONCLUSION

A new Delay-optimized 8-bit OR cell, power-optimized 8-bit Priority Encoder cell, and parallel priority look-ahead approach have been presented. A 64-bit Priority Encoder utilizing parallel look-ahead architecture pipelined structure has been developed on TSMC 130nm CMOS technology and 1.2V. Simulation have been done for the new design,and the conventional design after scaling the conventional design to be compatible with TSMC 130nm CMOS technology and simulate it again. Simulation results show that the new design is better than the conventional design as the critical path delay is reduced by a factor of 53%, and the total number of transistors is reduced from 647 to 599 which reduce the total chip area as shown in Table below, which indicates a fair comparison between the two designs at different frequency of operation.